IJLST

ISSN: 0974-5335
IJLST (2010), 3(2):16-28

Molecular cloning and structural characterization of catalytic domain of class III chitinase from *Tamarindus indica*.

Manali Datta and Shailly Tomar

Department of Biotechnology, Indian Institute of Technology Roorkee, Roorkee – 247667, Uttarakhand, INDIA
shailfbt@iitr.ernet.in

ABSTRACT:
Tamarindus indica, popularly known as Tamarind, which belongs to the fabaceae family is a commercially important plant. The tamarind seed kernel is a by-product of the pulp industries and could be a valuable source for protein extraction. An acidic class III chitinase, a member of the glycosyl hydrolase family 18, is the most predominant seed storage protein in tamarind seeds. Chitinases are industrially important enzymes that have wide range of applications such as crustacean chitin waste management. In an attempt to understand the structure function relationship, the catalytic domain of tamarind chitinase (cdCHT) was cloned and sequenced. Amino acid sequence deduced from nucleotide sequence indicates that cdCHT domain consists of 263 residues. Primary sequence analysis of cdCHT shows that it has high sequence homology with class III chitinases. Catalytic residues and substrate binding motifs were identified and found to be conserved in cdCHT. Tamarind chitinase has been reported to be a glycoprotein, and as expected three potential glycosylation sites were predicted in cdCHT primary sequence. The three-dimensional structure of cdCHT was constructed by homology modeling for structural characterization. Crystal structure of hevamine, a chitinase from *Hevea brasiliensis* with highest sequence homology with cdCHT was used as a template for model building. 3D model of cdCHT was energy minimized, loop regions were refined, and the final 3D structure was validated. Detail structure analysis and comparison revealed major differences in residues present in the loop regions involved in substrate binding. Thus, various potential substrates were docked into the final refined model of cdCHT. Docking studies with substrates indicate that the de-acetylated form of chitin would be a better substrate than chitin.

Key words: Chitinase, *Tamarindus indica*, catalytic domain, cloning, structure analysis

INTRODUCTION

Chitin is an unbranched polymer of N-acetyl β-D glucosamine, covalently linked as β 1, 4 – linkages and is present ubiquitously except in mammals. Chitin adds to a large amount of bio-waste as the natural decomposition of the polymer is time-consuming process.

A subclass of glycosyl hydrolases enables the degradation of chitin by hydrolysis of the β 1, 4 – glycosidic linkage and are classified as chitinases [EC 3.2.1.14]. Artificial topical application of chitinases or use of genetically engineered microbes has been considered as a solution for chitin waste treatment. Hence, the potential relevance of chitinases in bioremediation has increased. Chitinases occur extensively in a wide range of organisms including viruses, bacteria, fungi, insects, higher plants and some vertebrates [1, 2, 3].

Chitinases in plants are associated with a large number of physiological functions. They are mainly considered as pathogenesis-related (PR) proteins, since their activity is generally induced by microbial infections, wounding, elicitors such as salicylic acid, ethylene, auxins, cytokinins, heavy metal salts and by extreme soil and climatic conditions [4, 5, 6, 7] Chitinases have been established and identified as the marker enzyme in symbiosis [8]. Evidence for other physiological functions of chitinases in flowering, reproduction, germination and plant growth has also emerged. It has been implicated in embryogenesis [9], regulating the action of signal molecules [10, 11] and apoptosis [12].

Based on their primary structures, chitinases have been categorized into two major families of glycosyl hydrolases: family 18 and family 19, which are in turn subdivided into class I to VII [13]. Chitinases belonging to the family 18 includes the class III, class V and class VII chitinases. They also contain a group of inactive homologues termed as 'chito-lectins'. The mechanism of substrate catalysis involves a 'double displacement inverting' reaction for the chitinases belonging to class 18.

Chitinases of the family 19 encompasses the chitinases of class I, II, IV and VI [14]. Chitinases of this family have been found predominant in plants with recent findings in bacteria [15, 16]. Family 19 chitinases operate by an inverting mechanism which involves a direct attack of a nucleophilic water molecule on the sugar anomeric carbon and producing an α-anomeric product [17].

Family 18 chitinases have been isolated from diverse sources with equally diverse substrate preference [18]. The basis for this difference lies in the three dimensional arrangement of the amino acid moieties on the surface and active site of the enzyme. Thus, it becomes important to study the 3-D structure of chitinase to perform an interpretation of the structure

function relationship. Seeds of the plant have become the major source for purification of chitinases example being *H. vulgare* [19], *G. max* [20], *S. bicolor* [21], *P. glaucum* [22], *O. sativa* [23], *A. pavonina* [24] and *S. cereale* [25]. A member of the Leguminosae or Fabaceae family is *Tamarindus indica* (Tamarind). Tamarind is popular in India as a condiment. Technology has now been utilized to manufacture pectin, tartarates and alcohol from its pulp. The pulp of tamarind has medicinal virtues and is used in treating anorexia nervosa, as purgative, as an anti-helminthic and as antiseptics. The seed has found its applicability as textile thickener, in textile sizing of jute and cotton and in creaming rubber latex [26, 27, 28]. The tamarind fruit pulp industries produce a large amount of tamarind seed kernel as the by product. Thus, the kernel becomes an economic source for isolation of seed proteins such as chitinase or protease inhibitor having commercial and practical applications [29, 30].

In this study, we describe the molecular cloning and sequencing of the catalytic domain of chitinases (cdCHT) from *T. indica*. The deduced amino acid sequence of cdCHT was used and a theoretical 3D model of tamarind chitinase was predicted for structure function analysis. To our knowledge, this study is the first report on molecular cloning and 3D structure characterization of a protein from the plant species *T. indica*.

MATERIALS AND METHODS

Materials

Plant materials and chemicals.

RNase A, M-MulV reverse transcriptase, Taq polymerase, *Nde*I, *Xho*I, DNA ligase, PRnasin were obtained from Bangalore GeNei [Bangalore, India]. pET41b vector system was obtained from EMD4Biosciences, U.S.A. All other chemicals were obtained from Himedia, Mumbai. QIAquick gel extraction kit was obtained from Qiagen Inc, Valencia, CA. Sequencing of genes was done at Ocimum Biosolutions, Hyderabad, India. Diethylpyrocarbonate (DEPC) was obtained from Sigma Aldrich, U.S.A.

Hardware and software

Sequence alignment and presentation was done using CLUSTAL W2 [31] and ESPript [32] respectively. Automated homology modeling was done using MODELLER 9v7 (http://salilab.org). Model validation was done by PROCHECK [33], PROSA [34], ERRAT [35] and Verify-3D [36]. Energy minimization was done using Swiss PDB viewer (http://www.expasy.org/spdbv/). Docking studies were performed with Hex 5.0 [37]. Visualization of theoretical model was done using Pymol [38]. Homology modeling, docking and simulations work was performed in Red Hat Enterprise Linux 5 operation system (Red Hat Inc. Raleigh, NC) on a Dell Precision T5400 workstation.

Methods

Isolation of genomic DNA and total RNA

Freshly collected seeds were utilized as a source for RNA extraction. For total RNA isolation, tamarind pods were collected locally and surface sterilized using 70% ethanol and rinsed with autoclaved water. Manual crushing of the pod in liquid nitrogen was done and seeds thus obtained were stored in liquid nitrogen. Total RNA was isolated using the urea-lithium chloride method as previously described [39] with minor modifications. Briefly, 5 mg seeds were homogenized in liquid nitrogen and immediately transferred to 20 ml of lysis buffer [Urea 8M; lithium chloride 3M]. This was followed by overnight incubation at 4 ºC and centrifugation at 5000 x g for 45 minutes. The pellet was resuspended in resuspension buffer (0.5% SDS; 0.2M NaCl; 25 mM EDTA; 10 mM Tris pH 7.5 and 4 % Polyvinyl Pyrollidone) followed by a series of washes using phenol, phenol:chloroform and chloroform:IMA extractions. The supernatant was used further for final precipitation of RNA was performed with 2 % potassium acetate by overnight incubation in 4 ºC. Non-denaturing RNA gel was run to check the integrity of the isolated RNA and quantification was done by absorbance at 260 nm.

For genomic DNA isolation, freshly collected tamarind leaves were obtained locally and stored in liquid nitrogen. Genomic DNA was extracted from 1 gm of leaves by cetyl trimethylammonium bromide (CTAB) method [40]. Precipitated DNA was solubilised in 1X TE buffer pH 8 and subjected to RNase treatment (100 µg/µl) at 37 ºC for 2 hrs. The yield of DNA per gram of leaf tissue extracted was measured at 260 nm using a UV/VIS Spectrophotometer (Perkin Elmer, U.S.A). The qualitative estimation of DNA was performed by calculating the ratio of absorbance at 260/ 280 nm and by running a 0.8% TAE agarose gel.

Protein alignment and preparation of the degenerate primers

Chitinase from tamarind seeds has been purified and the N-terminal amino acid sequence has been reported for purified protein [29, 30]. Database similarity searches were performed for the reported N-terminal tamarind chitinase sequence using the BLAST tool from National Centre of Biotechnology Information (NCBI) website. Complete amino acid sequences of the homologous proteins from the non–redundant protein database were retrieved. Homologous sequences from the leguminosae family were selected and subjected to sequence comparison. Multiple sequence alignment of the selected homologus sequence was performed using the Clustal W2 software (http://www.ebi.ac.uk). Aligned sequences were analyzed and conserved amino acid sequence

motifs near the N-terminus and the C-terminus of class III chitinases were identified. For amplification of cdCHT, a pair of sense and antisense oligonucleotide primers complementary to these conserved motifs was designed. The sense primer,F1:5'-TATTGGGGCCAAAACGGY(C/T)A AY(C/T)GA-3' and the antisense primer, R1: 5'-CCAAAGAACCATAACACCACCATA-3' were used for the PCR reaction. For cloning purpose, the same pair of oligonucleotide primer containing *Nde*I and *Xho*I restriction enzymes sites: the sense primer, F2 5'-TAAGCGGCATATGTATT GGGGCCAAAACGGYAAYGA-3' (*Nde*I site underlined) and the antisense primer R2 5'-ATAATCCTCGAGCCAAAGAACCATAACACCAC CATA-3' (*Xho*I site underlined) were used.

RT-PCR and cDNA synthesis

Two micrograms of total RNA was hybridized with 50 pmol of oligo $d(T)_{18}$ primers and RNAs were reverse transcribed in 20 µl reaction volume containing 50 mM Tris-HCl pH 8.5, 8 mM $MgCl_2$, 30 mM KCl, 1 mM DTT, 1 mM of dNTPs, 20 U of PRNasin and100 U of M-Mulv reverse transcriptase. The RT cycle comprised of incubation of the reaction mix at 25 ºC for 5 min followed by a cycle of 37 ºC for 60 min in an Eppendorf thermal cycler. The RT reaction was terminated with a cycle of 5 min at 95 ºC. 3 µl of RT reaction product was utilized as a template for the synthesis of cDNA by 30 cycles of polymerase chain reaction [PCR]. cDNA was prepared in 50 µl total volume containing 5 µl of 10X reaction buffer, 2 µl of 20 mM dNTP, 1 mM primers (F1 & R1) and 3 U of Taq polymerase. PCR amplifications was performed in a thermocycler (Eppendorf AG Hamburg, Germany) with an initial denaturing step of 94 ºC for 5 min, followed by 30 amplification cycles of 94 ºC for 60 sec, 51 ºC for 60 sec, 72 ºC for 60 sec and a final extension cycle at 72 ºC for 5 min. PCR products were electrophoresed on a 1% TBE agarose gel, stained with ethidium bromide (EtBr) and visualized on a UV transilluminator.

Amplification of genomic DNA of cdCHT

Amplification of cdCHT genomic DNA carried out using the sense (F1) and the antisense (R1) primers. The same PCR program cycle was utilized for amplification of cdCHT using 3 µl of genomic DNA as template with some minor modifications. The PCR product was analyzed by electrophoresis on a 1 % agarose gel stained with EtBr.

Cloning and sequencing of cdCHT

cDNA and genomic DNA of cdCHT were used as template in a PCR reaction containing F2 and R2 primers having appropriate restriction enzyme for cloning these DNA fragments in pET41b vector. PCR products were gel extracted and purified using QIAquick gel extraction kit (Qiagen, U.S.A). Gel purified PCR products and pET41b vector were digested with *Nde*I and *Xho*I at 37 °C for 1 hr. The restriction enzyme digested PCR products and vector were again purified using QIAquick gel extraction kit. PCR products were ligated into pET41b using 0.5 µl T4 DNA ligase (Bangalore Genei, India) by incubating the ligation mixture overnight at 15 ºC. The ligation mixture was directly used for transformation of $CaCl_2$ competent DH5α cells by the heat shock method. Individual colonies were picked, grown overnight at 37 °C in LB broth containing kanamycin (80 µg/ml) for plasmid isolation. Recombinant plasmids were purified from overnight culture using plasmid isolation mini-prep kit (Qiagen, Inc. Valencia, CA). Isolated plasmids were digested using *Nde*I and *Xho*I restriction enzymes and also used as templates in PCR reactions to confirm the size of the inserts. Recombinant pET-CHTcDNA and pET-CHTgenomic plasmids were sequenced in both the direction by automated fluorescent sequencing on ABI PRISM 377 sequencer available at South campus, University of Delhi, India using universal T7 primers.

Sequence analysis and phylogenetic tree construction

Sequence identity of cloned cdCHT was verified by doing homology searches using the nucleotide Basic Local Alignment Search Tool (Blastn) algorithm. Sequence analysis tools of the ExPASy Server were used for processing nucleotide sequence of cdCHT to deduce the amino acid sequence. The obtained primary amino acid sequence was used for identification of potential glycosylation sites using the NetNGlyc 1.0 server [41]. Tentative phosphorylation signatures were also assigned using the NetPhos 2.0 server [42]. Blastp program was run with BLOSUM as a scoring matrix with a gap opening penalty of 11 and gap extension penalty of 1 to obtain homologues from the non redundant database. Computer assisted sequence alignment was accomplished by using the Clustal W2 program from EBI server and visually presented using EsPript. The resulting alignment was scrutinized using a maximum likelihood method and the TREEPLOT (www.bioinformatics.nl/tools/plottree) server was used to create the phylogenetic tree.

Homology modeling and validation of 3-D structure

Homology modeling for catalytic domain of chitinase from tamarind was performed in the following sequential steps: template selection from Protein Data Bank (PDB), sequence-template alignment, model building, model refinement and validation [43]. Template search for cdCHT was done using NCBI BLAST search tool against PDB database. The highest scoring hit was found to be with hevamine, a bifunctional chitinase/lysozyme from *Hevea brasiliensis* (PDB ID: 2HVM) and hence considering the favourable statistics [lowest E-value and highest identity], 2HVM was selected as the template for

ISSN: 0974-5335
IJLST (2010), 3(2):16-28

homology modeling. Modeller9v7 was used for generating three dimensional model of cdCHT. Fifteen preliminary models were generated which were ranked based on their negative DOPE scores. Five sets of models having lowest DOPE scores were selected and stereo-chemical quality of each was assessed by PROCHECK.

Energy minimization of the selected model was performed using Swiss-Pdb Viewer 4.01 (http://www.expasy.org/spdbv/). SPDBV implements GROMOS43B1 force field to compute energy and to execute energy minimization. Following energy minimization PROCHECK analysis for the obtained model was again done to check the favourable statistics.

The statistics for the favoured amino acid residue in Ramachandran plot, its accuracy and G- factor were considered for generation of the best model. The loopmodel tool of MODELLER was utilized for relieving steric clashes and improper contacts. Model with the least number of residues in the disallowed region was further refined in an iterative fashion till most of the amino acids were below 95% cut-off value in ERRAT plot. The refined model was further validated by VERIFY-3D of SAVES server (http://nihserver.mbi.ucla.edu/SAVES/). ProSA was used to evaluate the generated 3D structure model of protein for possible errors. The accuracy of the model generated was confirmed by calculating the root mean square deviation (RMSD) between the main chain of model (cdCHT) and template (2HVM) by superimposition using Pymol (http://www.pymol.org/).

Docking studies with potential substrates

The potential substrates for CHT were obtained in PDB format using the PRODRG server [44]. Docking experiments were performed using the program HEX.5 [37] which employs Spherical Polar Fourier (SPF) correlations in conjunction with a soft molecular mechanics potential function, thus improving the rank obtained for low RMS docking orientations. Relative stabilities were evaluated on the basis of free energy calculations done.

Nucleotide sequence and protein structure accession code

The cdCHT sequence has been deposited in GenBank database under accession number HM222538 and the coordinates for predicted 3D model have been submitted to PMDB database (http://mi.caspur.it/PMDB/; PMDB identifier no. PM0076336).

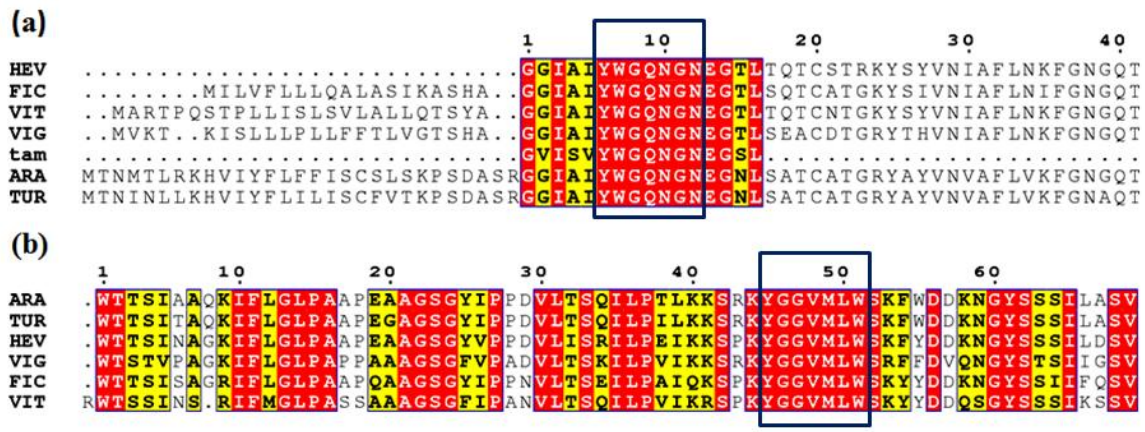

Figure 1. **[a]** N-Terminal and **[b]** C- terminal sequence alignment of the CHT with other chitinases retrieved from NCBI. Regions selected for primers are enclosed in box. VIG: *V angularis;* TUR: *T. glabra;* OLI: *O. pumila;* ARA: *A. thaliana*; HEV: *H. brasienlisis*; FIC: *F.awkeotsang;* VIT: *V. vinifera* TAMCHT: *T. indica*. Figure prepared using ESPript [http://espript.ibcp.fr/].

RESULTS AND DISCUSSION

Molecular cloning and sequencing of cdCHT

The N-terminal amino acid sequence of tamarind chitinase seed protein was used to search the NCBI protein database for homologus sequences. More than 25 sequences were positive hit with the query sequence and all belonged to a class III chitinases. Class III chitinase sequences from plant sources were selected and their primary sequences were retrieved and multiple sequence alignment was done using Clustal W2 program. A conserved amino acid sequence YWGQNG was identified near the N-terminus in the selected class III chitinase sequences (Fig.1a). Another

conserved peptide region YGGVMLW was identified near the C-terminus end (Fig. 1b). These highly conserved peptide regions were selected and sequence specific primer pair were designed for PCR amplification of catalytic domain of tamarind chitinase.

For isolation of total RNA, seeds were collected at different stages of maturation to check the expression of protein (data not shown). Forty days matured seeds were finally selected for RNA extraction by Urea-Lithium chloride method followed by selective precipitation of RNA using by potassium acetate. Two distinct bands corresponding to 28S and 18S rRNA bands were observed on 1.5% agarose gel in the isolated RNA sample. First strand cDNA synthesis by reverse transcription was done using oligo $d(T)_{18}$. Degenerate primers (F1 and R1) were used to amplify the target cdCHT DNA fragment using RT product as template. A PCR product of the expected size i.e. approximately 700 bps was amplified from cDNAs prepared from total RNA extracted from seeds. PCR product of the same size was also amplified from genomic DNA of *T. indica* indicating the absence of introns. For cloning and sequencing of cdCHT, F2 and R2 primers containing RE sites were used in a PCR reaction containing the amplified genomic DNA and cDNA as template. PCR products amplified using F2 and R2 were gel extracted and purified using the QIAquick gel extraction kit. Both the purified PCR products were subcloned into the *Nde*I and *Xho*I sites of pET41b vector. The ligation mixtures were used for transformation of DH5α cells; recombinant plasmids were isolated from overnight cultures and restriction enzyme digested to confirm the size of inserts. Both the recombinant pET-CHT cDNA and pET-CHT genomic plasmids were sequenced. Sequencing of cloned PCR products confirmed that the PCR product of cDNAs and genomic DNA was similar and the absence of introns was also confirmed.

Sequence and phylogenetic analysis

Sequence data obtained was analyzed using the Chromas lite software (www.technelysium.com.au). The obtained DNA sequence was translated and the amino acid was obtained.

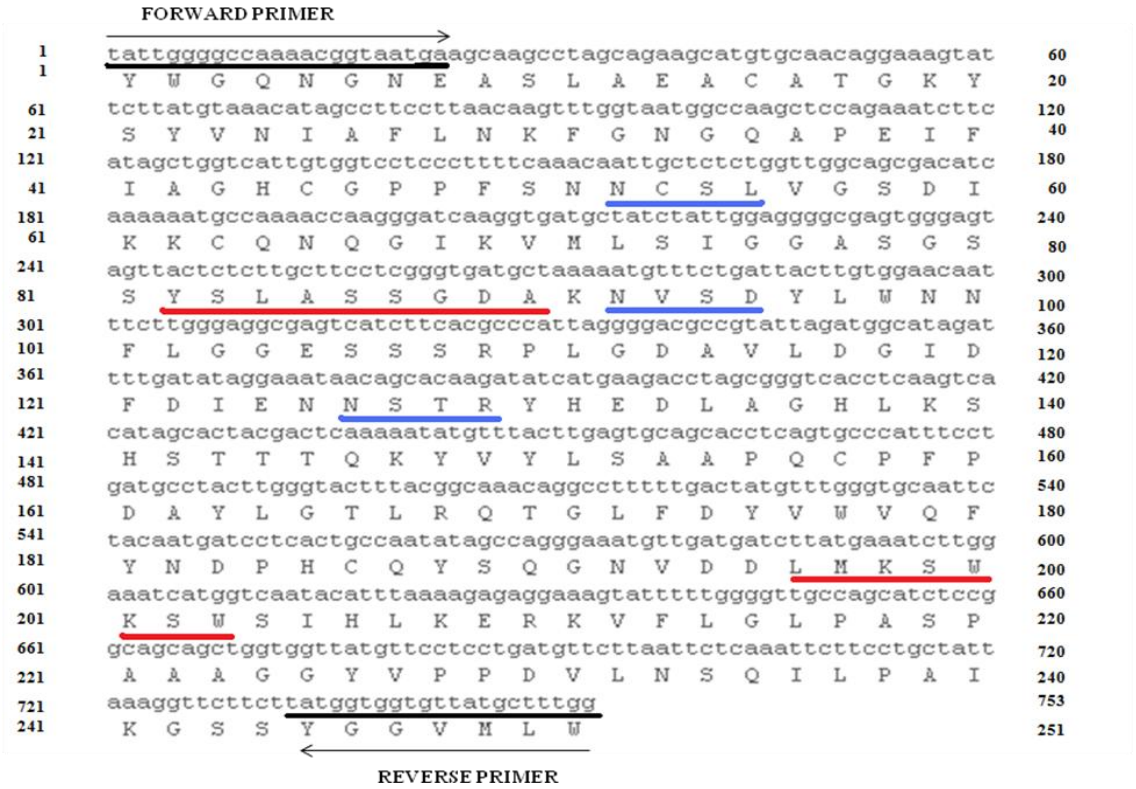

Figure 2. Nucleotide sequence of *Tamarindus indica* cdCHT. The deduced amino acid sequence is given in the one-letter code below the corresponding nucleotide sequence. Forward and reverse primers are marked by arrow in black. Potential glycosylation [blue] and phosphorylation [red] residues are marked. The nucleotide sequence of cdCHT has been deposited in GenBank under the accession number HM222538.

ISSN: 0974-5335
IJLST (2010), 3(2):16-28

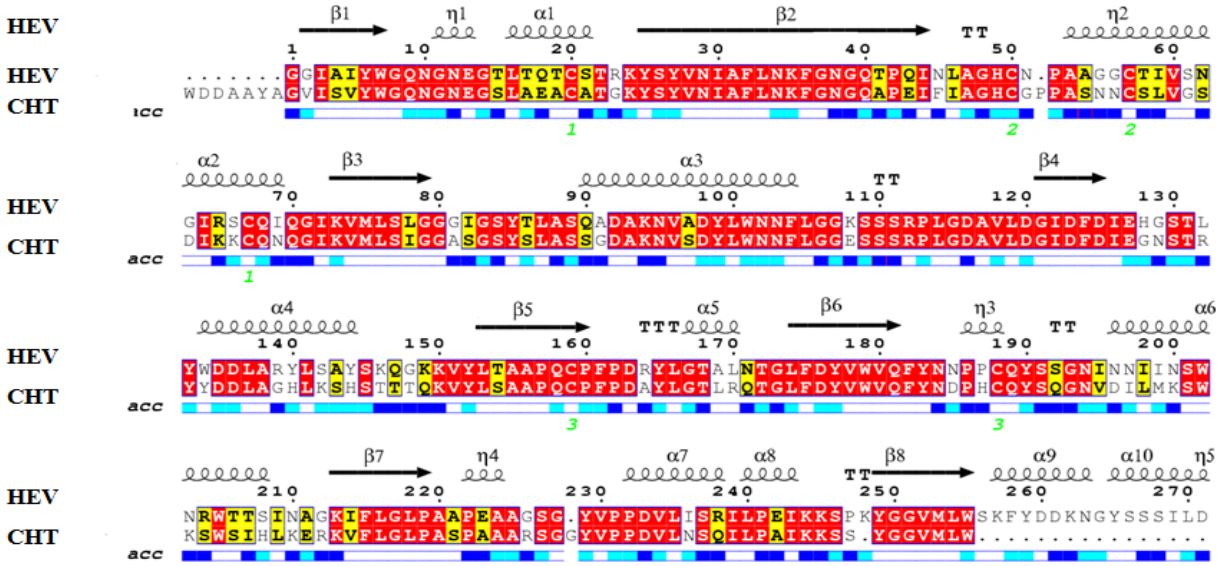

Figure 3. Secondary structure assignment to cdCHT based on the hevamine crystal structure. The solvent accessibility is indicated in blue highlights. Figure prepared using ESPript [http://espript.ibcp.fr/].

The deduced primary sequence of cdCHT consisting of 251 amino acid residues had highest identity of 67% with hevamine from *Hevea brasienlisis*. Since chitinase is the major glycoprotein present in tamarind seeds, the potential N-glycosylation sites was determined using NetNGlyc 1.0 server [29, 30].

Three regions in the primary sequence of cdCHT were identified as glycosylation motifs at position $_{54}$NCSL$_{58}$, $_{92}$NVSD$_{95}$ and $_{126}$NSTR$_{129}$ with glycosylation potential of 0.6198, 0.6222 and 0.7042 respectively (Fig. 2). Evaluating the glycosylation motifs of cdCHT with other members of the class III chitinases, it was observed that there was a characteristic absence of these motifs in the other members of this family.

Chitinases from some plant sources have also been implicated in signal transduction and apoptosis [45]. So the possibility of the protein being phosphorylated also seemed promising. Computational prediction of phosphorylation sites was accomplished by NetPhos 2.0 server. Two sites with the possibility of being phosphorylated were identified at positions $_{82}$YSLASSGDA$_{90}$ (0.99) and $_{196}$LMKSWKSW$_{203}$ (0.97) (Fig 2).

Pairwise sequence alignment of cdCHT with hevamine (Fig. 3) was performed to compare the secondary structural elements. The overall comparison of the sequence gave the credible indication of the characteristic Rossmann fold being conserved in the cdCHT. Additionally, the key residues which are universally conserved within the glycosyl hydrolase 18 family were also identified and found to be conserved in cdCHT (Fig 4). There were substitutions and insertions in the loop region thereby affecting their flexibility (insertion at position 59 with a proline residue; substitution with a histidine at position 195).

The sequence has six conserved cysteine residues forming three intra-chain disulphide bonds at positions Cys27-Cys75, Cys 57-Cys65 and Cys 167-Cys201corresponding to conserved structure of Hevamine (2HVM). Phylogenetic tree for cdCHT was constructed based on neighbor joining distance and maximum likelihood methods (Fig. 5) using the TREEPLOT server.

The divergence of hevamine (HEV), PPL2, a lectin from *Parkia platycephala* (PAR) and cdCHT (TAMCHT) from a same node indicated they have a common evolutionary history. PPL2 has formed an outgroup indicating greater evolutionary distance from cdCHT, which was evident in the sequential dissimilarity observed between the two proteins.

Homology modeling and structure validation of CHT from T. Indica

Crystal structure of only two plant class III chitinases has been determined. Based on these crystal structures, the catalytic mechanism of acidic class III chitinases

has been analyzed in details but the substrate specificity has not been addressed.

The difference in substrate specificity and affinity of chitinases is determined by the amino acid compositions at substrate recognition site and the 3D conformation of substrate binding site [46, 47]. In this report, 3D structure prediction and substrate docking studies to investigate substrate specificity, amino acid sequence and structure analysis was carried out. Crystal structure of hevamine (PDB ID: 2HVM) having maximum sequence identity of 67% with cdCHT was selected as a template for construction of 3D homology model of tamarind chitinase.

Preliminary 3D models that were generated using MODELLER9v7 were validated using PROCHECK to evaluate the stereo-chemical quality of the model. Preliminary model with most favourable statistics was selected and subjected to refinement, loop modeling and energy minimization. PROCHECK, Verify_3D and ERRAT plot were used for determining the stereo-chemical parameters of the energy minimized model of cdCHT.

Ramachandran plot generated by PROCHECK for the final refined and energy minimized 3D model of cdCHT shows that 98.6% residues are present in the allowed region, 0.9% in the generously allowed region and 0.5% in the disallowed region. ProSA plot also indicates that the generated 3D model of cdCHT is in agreement with hevamine (PDB ID: 2HVM), the template selected for homology modelling (Fig. 6).

The overall interaction energy of the model was −9.19 kcal/mol, which was found to be quite comparable to the template 2HVM (−8.84 kcal/mol). Verify_3D data server indicates that 99.22 % residues of the model have a favourable 3D-1D score. Structural comparison of cdCHT 3D homology model with the crystal structure of selected template was done to further validate the predictions. The RMSD of Cα trace between hevamine from *H. brasiliensis* and cdCHT from *T. indica* structures was 0.149 Å RMSD.

The confirmatory data obtained from the geometry and energy profiles indicate that the 3D model constructed for cdCHT of *T. indica* by comparative modeling is reliable for detail structural analysis.

Overall structure and active site geometry

Topology of the cdCHT is conserved in having a TIM $(\alpha/\beta)_8$ barrel domain structure characteristic of the class 18 hydrolases (Fig. 7a and Fig 7b). X-ray studies have shown that enzymes of GH18 family having chitinases activity have a conserved Asp125, Glu127 and Tyr183 amino acids (Hevamine numbering) in the active site. The molecular mechanism involves Glu127, which is considered as the proton donor to the glycosidic bond and Asp125 and Tyr183 that help in stabilizing the intermediate.

The conserved glutamate present at the C-terminus of the β4 helix is placed at position 122 in cdCHT. The aspartate residue which has a role in stabilizing the oxazolinium intermediate of the chitin-oligosaccharide is placed optimally at position 124.

These two residues tend to form an important part of the active site residues. Glu122 is fixed in this position by a hydrogen bond with Asp124 in cdCHT 3D structure. The other potential aspartates which can contribute to stabilization are placed far away from the active site and hence may not play important role in stabilizing substrate binding.

Loops connecting the carboxy terminus of α-helices with the amino terminus of β-strands in cdCHT are generally 4-7 amino acids long. There is 7-12amino acid residues length variation in loops that connect the carboxy terminus of β-strands with the amino terminus of α-helices. Apart from the characteristic TIM barrel β strands, there exists an extra short β-strand in cdCHT 3D structure after β2 consisting of three residues [Glu50-Phe52].

The consensus motifs characteristic of family 18 glycosyl hydrolases is conserved in spatially in β3 ($_{89}$KVI/L/MLSI/LGG$_{96}$) and β4 ($_{119}$DXXDXDXE$_{126}$) strands. Although the β4 strand is totally conserved with respect to hevamine (PDB ID: 2HVM) and PPL2, a chitinase from *Parkia platycephala* (PDBID: 2GSJ), the β3 strand has an amino acid substitution in case of cdCHT.

The substitution is leucine (position 78 in 2HVM) to isoleucine (position 74 in cdCHT) and amino acid residue at this position participates in substrate binding. Therefore, this indicates that hydrophobic interactions in this region of cdCHT protein could also be crucial for substrate binding and stabilization. 3D structural comparison of cdCHT with class III acidic chitinase (hevamine and PPL2) shows variation in the loops regions.

Most of the loop regions were observed to be superimposed with hevamine but high degree of variation was noticed when the structure of cdCHT was superimposed on PPL2 from *Parkia platycephala* (2GSJ) (data not shown). The comparative sequence and structure data obtained for 2GSJ and cdCHT were in complete agreement with the phylogenetic data and revealed the cause of distance responsible in the evolutionary history.

ISSN: 0974-5335
IJLST (2010), 3(2):16-28

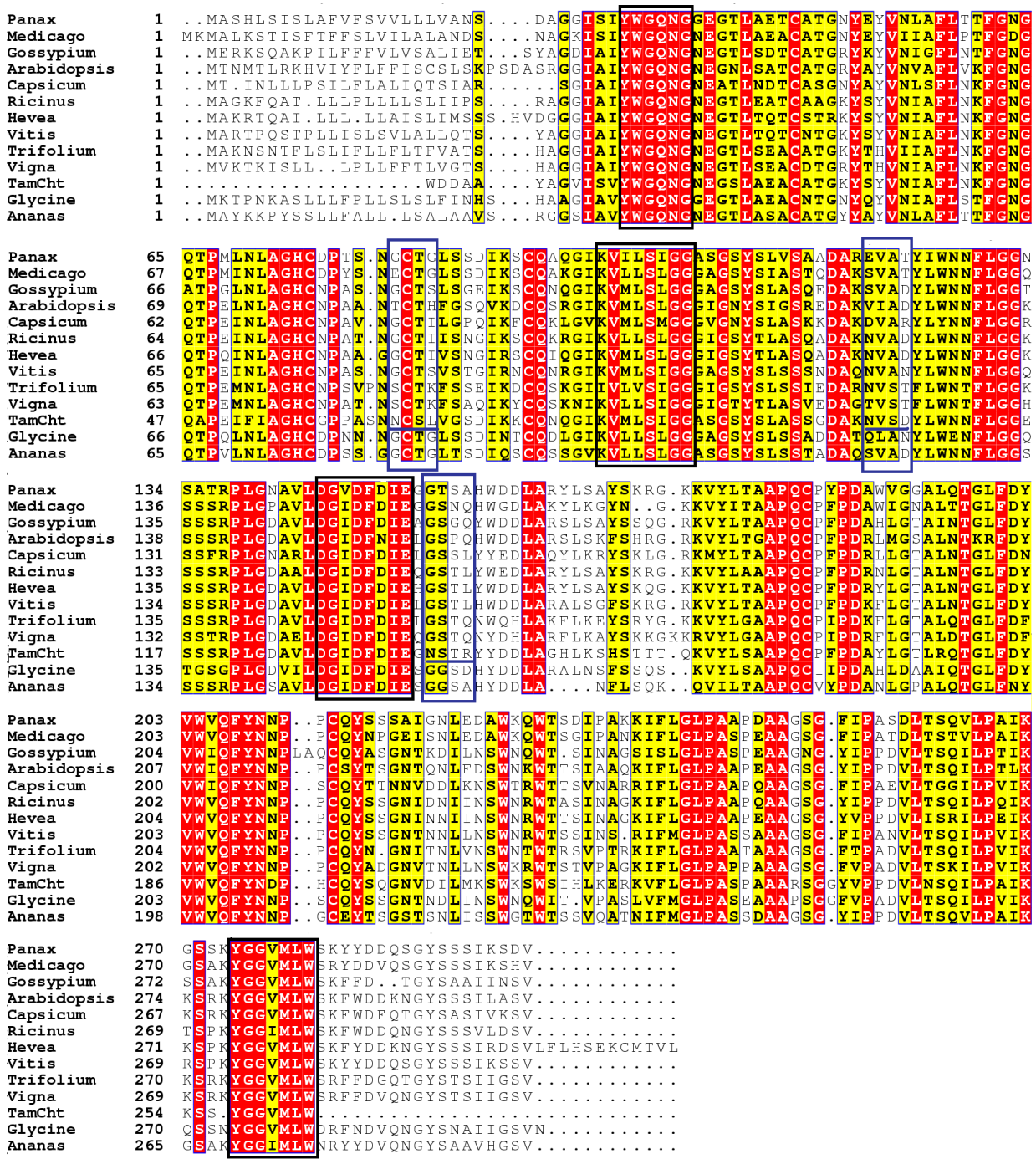

Figure 4: Multiple sequence alignment with similar sequences retrieved from NCBI protein database. *Panax ginseng* [gb|ABF82271.1|]., *Medicago truncatula* [gb|AAQ21404.1|]., Gossypium hirsutum [gb|ABN03967.1|]., *Arabidopsis thaliana* [dbj|BAA21874.1|]., *Capsicum annuum* [gb|AAN37389.1]., *Ricinus communis* [XP_002511935]., *Hevea brasiliensis* [emb|CAA09110.1]., *Vitis vinifera* [gb|ACH54087.1]., *Trifolium repens* [emb|CAB65476.2], *Vigna unguiculata* [emb|CAA61279.1], *Glycine max* [dbj|BAA77675.1], *Ananas comosus* [dbj|BAG38685.1]. Conserved domains are highlighted by a black box. The blue highlighted box indicates the potential regions of glycosylation in cdCHT. Figure prepared using ESPript [http://espript.ibcp.fr/].

ISSN: 0974-5335
IJLST (2010), 3(2):16-28

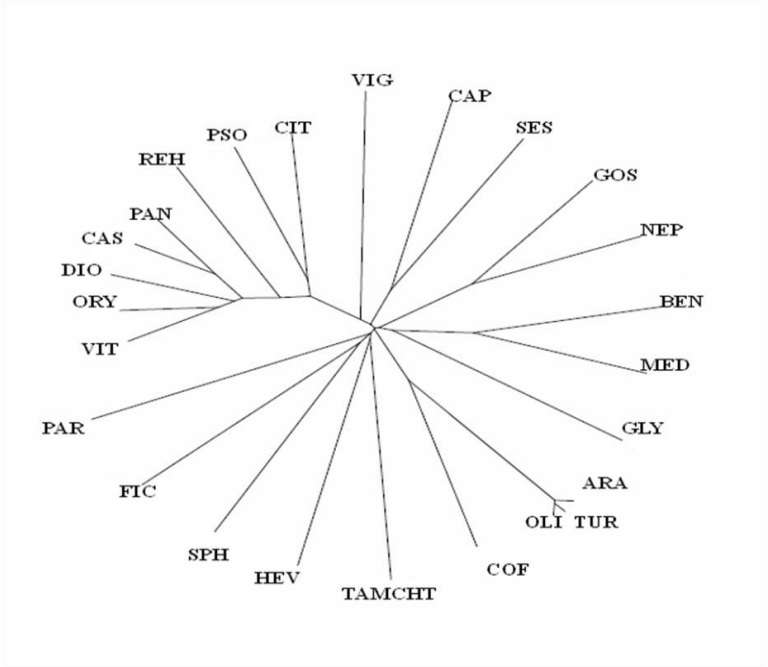

Figure 5. TREEPLOT made from the sequences retrieved from the NCBI protein database based on maximum likelihood method cluster analysis. **PAN:** *Panax ginseng* (gb|ABF82271.1|), **MED**: *Medicago truncatula* (gb|AAQ21404.1|)., **GOS:** *Gossypium hirsutum* (gb|ABN03967.1|)., **ARA**: *Arabidopsis thaliana* (dbj|BAA21874.1|)., **CAP**: *Capsicum annuum* (gb|AAN37389.1)., **HEV:** *Hevea brasiliensis* (emb|CAA09110.1)., **VIT:** *Vitis vinifera* (gb|ACH54087.1), **VIG**: *Vigna unguiculata* (emb|CAA61279.1)., **GLY:** *Glycine max* (dbj|BAA77675.1)., **REH:** *Rehmannia glutinosa* (gb| AAO47731|)., **DIO:** *Dioscorea oppositofolia* (BAC56863.1)., **TUR**: *Turritis glabra* dbj|BAA21876.1|; **NEP:** *Nepenthes rafflesiana* (ACU31856.)., **BEN:** *Benincasa hispida* **(**gb|AAD56239.1)., **OLI** *Olimarabidopsis cabulica* **(**BAC11882.1)., **COF:** *Coffea arabica* **(**CAJ43737.1)., **SPH:** *Sphenostylis stenocarpa* **(**AAD27874.1)., **FIC:** *Ficus pumila* **(**AAQ07267.1)., **PAR:** *Parkia* p*latycephala* (2GSJ_A)., **ORI:** Oryza sativa Japonica (BAC55717.1)., **CAS:** *Casuarina glauca* (ABL74451.1**).,** **PSO:** *Psophocarpus tetragonolobus* **(**BAA08708.1)., **CIT:** *Citrullus lanatus* **(**ABA26457.1)., **SES:** *Sesbania rostrata* **(**CAA88593.1)

Docking of substrates

Docking studies using the final refined 3D model of tamarind chitinase were performed to elucidate the structural and functional relevance in terms of substrate binding and specificity. NAG polymers namely tri, tetra and penta (Fig. 8a) were successfully docked onto cdCHT 3D structure using HEX 5.0. Structures analysis of cdCHT substrate complexes was done in details to identify the substrate binding residues and molecular interactions playing crucial role in substrate specificity. In case of penta-NAG [N-acetyl glucosamine], the amino acid residues lining the substrate binding cavity are Tyr 13, Gln 16, Asn 17, Gly 18, Asn 19, Phe 39, Asn 41, Phe 52, Ala 54, Gly 55, Ala 80, Gly 86, Ala 88, Ser 89, Ser 90, Asp133, Glu135, Ala 232 and Trp 263. However, it was observed that the interactions of N-acetyl groups were always hindered due to sterical clashes with the side chains of some residues in cdCHT. For example, when trimer of N-acetyl glucosamine (tri-NAG) was docked and analyzed, the N-acetyl group showed unfavorable interactions, particularly with the amino acid residue at position 52 (Phe 52). Asn 45 (incase of 2HVM) has been substituted by Phe 52 in cdCHT which causes massive amount of stereo conflict in the substrate binding site. Similarly the polar contact with residue Asn 41 found in hevamine was not possible as the rotamer present did not give favorable orientation to enable interaction with N-acetyl group (Fig. 8b). This observation made by molecular docking of substrates suggested that cdCHT might have more affinity for the de-acetylated form of chitin. Further, to investigate this, docking studies were carried out using the deacetylated form of tri-NAG (dNAG) on the modeled structure of cdCHT and the relative stabilities was evaluated by molecular dynamics using free energy simulations. With the deacetylated form of substrate, no steric clashes were identified. The lead molecule is the one having maximum interaction with high negative E-

value. Based on the E- value obtained for dNAG (396.01 as compared to -334.28 of tri-NAG), it was confirmed that chitosan (deacetylated chtin) could be a better substrate for *T. indica* chitinase. Rao and Gowda, 2008 have reported that the class III chitinases, which was isolated from tamarind seeds, has low chitinase activity. We suspect that the structural incompatibility of chitin with cdCHT could be an explanation for low enzymatic activity of tamarind chitinase. In future, the role of Phe52 in substrate specificity can further be elucidated and confirmed by site-directed mutagenesis experiments

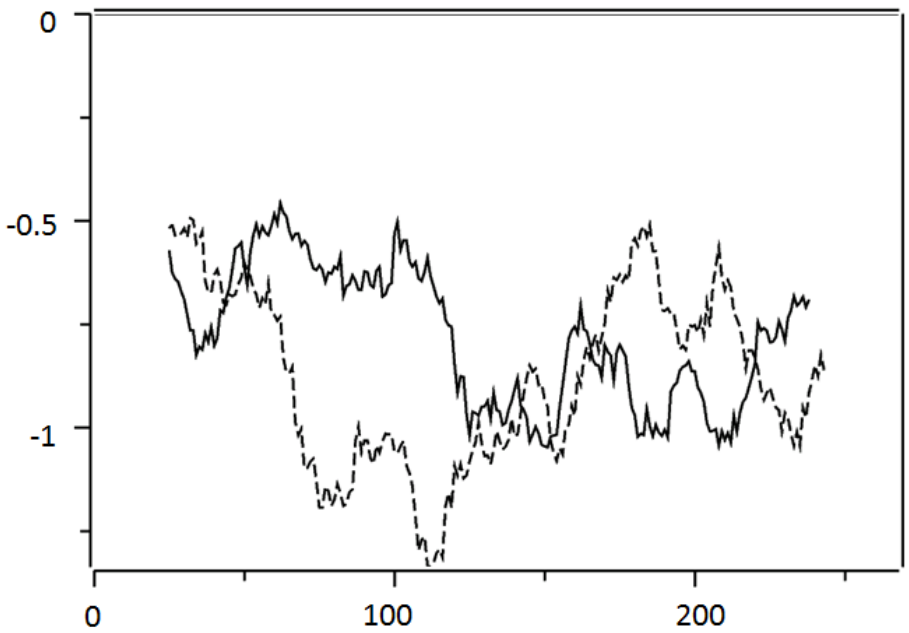

Figure 6. ProSA energy plot for *Tamarindus indica* cdCHT model. The dotted lines indicate hevamine template and solid line indicates cdCHTmodel.

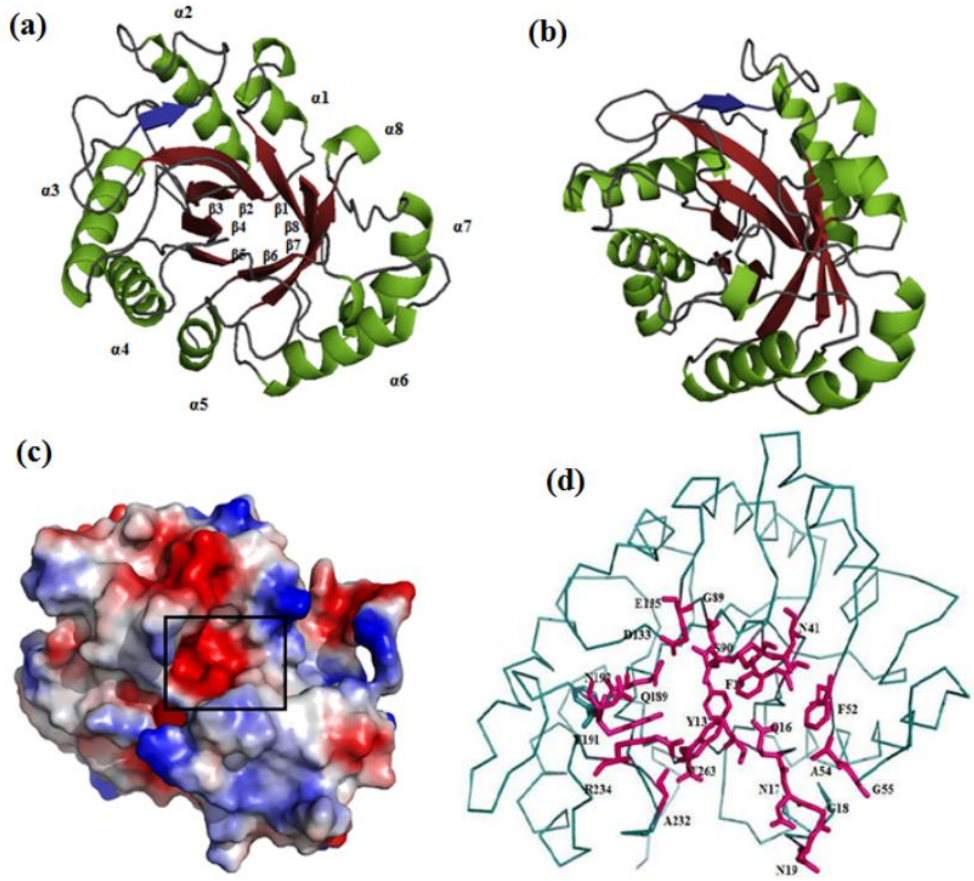

Figure 7. [a] Top view of cdCHT 3D model generated by MODELLER. [**b**] Side view of cdCHT 3D model indicating the characteristic α/β_8 barrel. **[c]** Electrostatic surface potential for cdCHT with marked active site; the polar residues are shown in red and blue and hydrophobic residues are shown in white. **[d]** Ribbon diagram of cdCHT with residues [in stick representation] in the catalytic cleft marked in pink. 3D-model of cdCHT has been submitted to PMDB database [PM0076336]. Figures prepared using PyMOL [http://www.pymol.org].

IJLST

ISSN: 0974-5335
IJLST (2010), 3(2):16-28

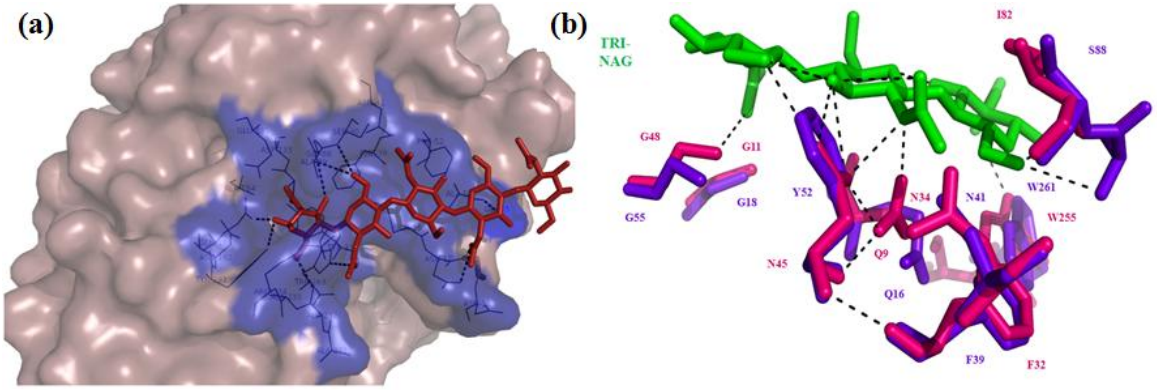

Figure 8 [**a**] Placement of penta-NAG in the active site of cdCHT with the active side residues Glu 135 [pink], Asp133 [yellow] and Tyr191 [blue] [**b**] Comparison of active site geometry of 2HVM [hevamine][pink] and cdCHT[violet] with chitotriose placed in the active site cleft. Figures prepared using PyMOL [http://www.pymol.org/].

CONCLUSIONS

T. indica chitinase from tamarind seeds has been cloned and sequenced. cdCHT DNA fragment encoding a stretch of 251 amino acids, was cloned both from cDNA and genomic DNA. Sequencing of the products revealed that both cDNA and genomic DNA sequences are similar and confirmed the absence of introns.

Primary sequence analysis confirmed that cdCHT belongs to a class III acidic chitinases. Homology model of cdCHT was constructed and substrate docking studies were performed. Sequence analysis and comparison showed that cdCHT primary sequence contains some distinct crucial residues in the substrate binding site that may play a role in substrate specificity and affinity. Comparative docking of tri-NAG (chitin) and deacetylated NAG (chitosan) as substrates confirmed the above observation. Tri-NAG on docking with cdCHT produced an energy value of -334.28. Slight modification of tri-NAG by converting it into de-acetylated form lowered the energy value to -396.01.

The binding site of the dNAG remained same as tri-NAG, which indicates that functional groups involved are same and only the stearic compatibility increases by substitution of dNAG. These results suggest that cdCHT could be a better chitosanase than chitinases [48]. Another anomality with cdCHT was that it had propensity to be glycosylated which is unusual for other chitinases as they the lack these motifs.

It has been shown that glycosylation results in the structural and functional diversification of a single protein to yield a glyco-isoforms [49]. Additionally, glycosylation adds to *in vivo* half life and bio-distribution [45, 50]. Majority of glycosylated proteins in seeds are lectins and play a role in seed survival. This means that tamarind chitinase being the most abundant glycoprotein in seed might have lectin-like properties and needs to be investigated in future. Some of the chitinases also have function in signal transduction.

Presence of potential phospholyration sites in primary sequence of cdCHT points towards the possibility of its role as signaling molecule in seed germination. It has already been proved that a variant of chitinase, SL-CLP interacts with stabilin-1. Stabilin-1has been identified to be involved in adhesion and trans-migration in various cell types [51].

Probable association and interaction of cdCHT with a stabilin-like protein might shed unique capability of cdCHT in signal transduction. Comprehensive biochemical, biophysical and mutational studies are necessary and are being pursued to define multifunctional role of tamarind chitinase.

ACKNOWLEDGEMENT

This work was supported by financial aid from the Department of Science and Technology, New Delhi, India. Manali was supported by AICTE, New Delhi. We thank Dr. Pravindra Kumar and Dr. Ashwani Sharma for critical review of the manuscript. We also acknowledge Macromolecular Crystallographic Unit [MCU] at Indian Institute Technology, Roorkee (IITR) for providing cloning and computational facilities.

REFERENCES

[1] Gooday G. W. (1990) Inhibition of chitin metabolism, In *The Biochemistry of Cell Walls & Membranes in Fung*, pp, 61-79, Edited by P, J, Kuhn, A, P, J, Trinci, M, J, Jung, M, W, Goosey & L, G, Copping, Berlin: Springer.

[2] Cohen-Kupiec R., I. Chet (1998) The molecular biology of chitin digestion. *Curr, Opin, Biotechnol*, 9: 270-277.

[3] Felse P. A., T. Panda (1999) Studies on Applications of Chitin and its Derivatives. *Bioprocess Eng*, 20: 505-512.

[4] Henrissat B., G. J. Davies (1997) Structural & sequence-based classification of glycoside hydrolases. *Curr. Op. Struct. Biol*, 7: 637-644.

[5] Davies G., B. Henrissat (1995) Structures and mechanisms of glycosyl hydrolases. *Structure*, 3: 853-859.

[6] Ohno T., S. Armand, T. Hata, N. Nikaidou, B. Henrissat, M. Mitsutomi T. Watanabe (1996) A modular family 19 chitinase found in the prokaryotic organism *Streptomyces griseus* HUT 6037. *J Bacteriol*, 178: 5065-5070.

[7] Kong H., M. Shimosaka, Y. Ando, K. Nishiyama, T. Fujii, K. Miyashita (2001) Species- specific distribution of a modular family 19 chitinase gene in *Burkholderia gladioli*. *FEMS Micro, Ecology*, 37 135-141.

[8] Ueda M., T. Kojima, N.Yoshikawa, K. Araki, T. Kawaguchi, K. Miyatake, M. Arai, T. Fukamizo (2003) A novel type of family 19 chitinase from *Aeromonas* sp, No,10S-24, Cloning, sequence, expression, and the enzymatic properties. *Eur J Biochem*, 270:2513-2520.

[9] Fukamizo T., D. Koga, S. Goto (1995) Comparative biochemistry of chitinases – anomeric form of the reaction-products. *Biosci, Biotech, Biochem*, 59:311-313.

[10] Graham L. S., M. B. Sticklen (1994) Plant chitinases. *Can J Bot*, 72:1057-1083.

[11] Dana M. M., J. A. Pintor-Toro, B. Cubero (2006) Transgenic tobacco plants overexpressing chitinases of fungal origin show enhanced resistance to biotic and abiotic agents. *Plant Physio*, 142:722-730.

[12] Nakamura T., M. Ishikawa, H. Nakatani, A. Oda (2008) Characterization of Cold-Responsive Extracellular Chitinase in Bromegrass Cell Cultures and Its Relationship to Antifreeze Activity. *Plant Physiology*, 147:391-401.

[13] Schraudner M., D. Ernst, C. Langebartels, H. Sandermann (1992) Biochemical plant responses to ozone, III, Activation of the defense-related proteins β-1, 3-glucanase and chitinases in tobacco leaves. *Plant Physiol*, 99:1321-132.

[14] Bonanomi A., A Wiemken, T. Boller P. Salzer (2001) Local induction of a mycorrhiza-specific class III chitinase gene in cortical root cells of *Medicago truncatula* containing developing or mature arbuscules. *Plant Biology*, 3:194-199.

[15] De Jong A., J. Cordewener, F. LoSchiavo, M. Terzi, J. Vandekerckhove, A Van Kammen, S. De Vries (1992) A carrot somatic embryo mutant is rescued by chitinase. *Plant Cell*, 4:425-433.

[16] Goormachtig S., S. Lievens, W.Van de Velde, M. Van Montagu, M. Holsters (1998) Srchi13, a novel early nodulin from *Sesbania rostrata*, is related to acidic class III chitinases. *Plant Cell*, 10:905-912.

[17] Helleboid S., T. Hendriks, G. Bauw, D. Inze, J. Vasseur J. L. Hilbert, (2000) Three major somatic embryogenesis related proteins in *Cichorium* identified as PR proteins. *J, exp, Bot*, 51:1189-1200.

[18] Passarinho P. A., A. J. Van Hengel, P. F. Fransz, S. C. de Vries (2001) Expression pattern of the *Arabidopsis thaliana* AtEP3/AtchitIV endochitinase gene. *Planta*, 212: 556-567.

[19] Leah R., H. Tommerup, I. Svendsen J. Mundy (1991) Biochemical and molecular characterization of three barley seed proteins with antifungal activity. *J, Biol, Chem*, 66:1464-1573.

[20] Krishnaveni S., S. Muthukrishnan, G. H. Liang, G. Wilde, A. Manickam (1999) Induction of chitinase and β-1,3-glucanases in resistant and susceptible cultivars of sorghum in response to insect attack, fungal infection and wounding. *Plant Sci*, 144: 9-16.

[21] Yeboah N. A., M. Arahira, V. H. Nong, D. Zhang, K. Kadokura, A. Watanabe C. Fukazawa, (1998) A class III acidic endochitinase is specifically expressed in the developing seeds of soybean (*Glycine max* (L,) Merr,). *Plant mol biol*, 36: 407-15.

[22] Radhajeyalakshmi R., B. Meena, R. Thangavelu, S.D. Deborah, P. Vidhyasekara, R. Velazhahan. (2006) A 45-kDa chitinase purified from pearlmillet (*Pennisetum glaucum* (L,) R, Br,) shows antifungal activity. *J. Pl. Di. And Prot*, 107: 605-616.

[23] Baek J. H., B. K. Han, D. H. Jo (2001) Distribution of chitinase in rice (*Oryza sativa* L) seed and characterization of a hull specific chitinase. *J, Biochem, Mol, Biol*, 34:310-31.

[24] Santos I. S., A. E. Oliveira, M. D. Cunha, O. L. Machado, A. G. Neves-Ferreira, K. V. Fernandes, A. O. Carvalho, J. Perales, V. M. Gomes, (2007) Expression of chitinasesin *Adenanthera pavonina* seedlings. *Physio, Plant*, 131: 80-88.

[25] Taira T., T. Ohnuma, T. Yamagami, Y. Aso , M.Ishiguro (2002) Antifungal activity of rye (*Secale cereale*) seed chitinases: the different binding manner of class I and class II chitinases to fungal cell wall. *Biosci, Biotechnol, Biochem*, 66:970-977.

[26] Rao K H., N. Subramaniam (1984) Nitrogen solubility and functional properties of tamarind seed kernel proteins. In *Proceedings of the National Symposium on Protein Foods and Feeds*, A67-A87.

[27] Bal S., R.K. Mukherjee (1994) Functional and nutritional properties of Tamarind kernel protein. *Food Chem*, 49:1-9.

[28] Patil S. J., B.S. Nagagouder (1997) Industrial Products from *Tamarindus Indica*. *Proc, Nat. Sym, on Tamarindus indica L*, 151–155.

[29] Rao D. H., L.R. Gowda (2008) Abundant class III acidic chitinase homologue in tamarind (*Tamarindus indica*) seed serves as the major storage protein. *J Agric Food Chem,* 56:2175-82.

[30] Patil D. N., M. Datta, A. Chaudhary , A.K. Sharma, S. Tomar P. Kumar (2009) Isolation, purification, crystallization and preliminary crystallographic studies of chitinase from Tamarind seeds. *Acta cryst F,* 65:343-345.

[31] Larkin M. A., G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam , F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson, D.G. Higgins. (2007) ClustalW and ClustalX version 2. *Bioinformatics*, 23:2947-2948.

[32] Gouet P., E. Courcelle , D. I. Stuart F. Metoz (1999) ESPript: multiple sequence alignments in PostScript. *Bioinformatics*, 15:305-308.

[33] Laskowski R. A., M. W. MacArthur, D. S. Moss, J.M. Thornton (1993) PROCHECK – a program to check the stereochemical quality of protein structures. *J. Appl. Cryst*, 26:283-291.

[34] Wiederstein M., M. J. Sippl (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res*, 35:407-410.

[35] Colovos C., T.O. Yeates (1993) Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci*, 9:1511-1519.

[36] Eisenberg D., R. Lüthy, J.U. Bowie (1997) VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods Enzymol*, 277:396-404.

[37] Ritchie D. W., G. J. Kemp (2000) Protein docking using spherical polar Fourier correlations. *Proteins*, 39:178-194.

[38] DeLano W. L. (2002) The PyMOL Molecular Graphics System DeLano Scientific, San Carlos, CA, USA, (http://www.pymol.org).

[39] Tai H. H., C. Pelletier, T. Beardmore (2004) Total RNA Isolation from *Picea mariana* Dry Seed. *Plant Molecular Biology Reporter*, 22:93a-93e .

[40] Doyle J.J., Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue, Phytochem Bull 19:11-15.

[41] Julenius K., A. Molgaard , R. Gupta S. Brunak S (2005) Prediction, conservation analysis and structural characterization of mammalian mucin-type O-glycosylation sites. *Glycobiology*, 15:153-164.

[42] Martí-Renom M. A., A.C. Stuart, A. Fiser , R. Sánchez , F. Melo, A. Sali (2000) Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct*, 29: 291-325.

[43] Schuettelkopf A. W., D.M.F van Aalten (2004) PRODRG - a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crys D,* 60:1355-1363.

[44] Hartl D., C. H. He, B. C. Koller, A. Da Silva , Y. Kobayashi, C. G. Lee, R.A. Flavell J.A. Elias (2009) Acidic Mammalian Chitinase Regulates Epithelial Cell Apoptosis via a chitinolytic independent Mechanism. *J, Immunol*, 182:5098-5106.

[45] Li H, L. H. Greene, (2010) Sequence and structural analysis of the chitinase insertion domain reveals two conserved motifs involved in chitin-binding. *PLoS*, 5: 1-11.

[46] Waddling C. A., T. H. Plummer Jr, A. L. Tarentino, P. V. Roey (2000) Structural basis for the substrate specificity of endo-*a*-*N*-acetylglucosaminidase F3. B*iochemistry*, 39:7878-7885.

[47] Osswald W. F., J. P. Shapiro, H. Doostdar, R. E. McDonald, R. P. Niedz, C. J. Nairn, C. J. Hearn, R. T. Mayer, (1994) Identification and Characterization of Acidic Hydrolases with Chitinase and Chitosanase Activities from Sweet Orange Callus Tissue. *Plant and Cell Physio*, 35 811-820.

[48] Rasmussen J. R. (1992) Effect of glycosylation on protein function. *Current Opinion in structl Bio*, 2 682-686.

[49] Parekh R. B (1991) Effects of glycosylation on protein function. *Current Opinion in Struct Biol*, 1:750-754.

[50] Kzhyshkowska J., A. Gratchev, S. Goerdt (2007) Stabilin-1, a homeostatic scavenger receptor with multiple functions. *J. Cell. & Mol, Med*, 3:635-649.